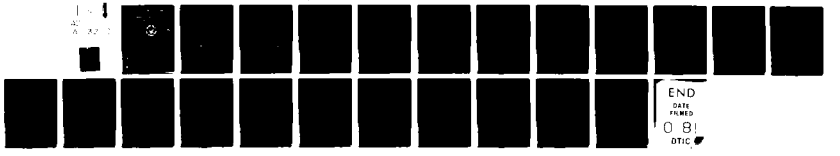
AD-A103 280 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 17/2
EXAMINATION OF VOICE RECOGNITION SYSTEM TO FUNCTION IN A BILING--ETC(U)
MAR 81 D E NEIL, T ANDREASON
UNCLASSIFIED NPS55-81-004 NL

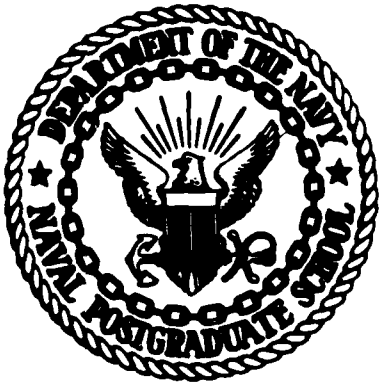


END
DATE FILMED
08
DTIC

(14) NPS55-81-004

NAVAL POSTGRADUATE SCHOOL

Monterey, California



(1) Technical rept.

(6) EXAMINATION OF VOICE RECOGNITION SYSTEM TO FUNCTION IN A BILINGUAL MODE

by

(10) D. E. Neil
T. Andreason

(11) Mar[redacted]81

(12) 25

Approved for public release; distribution unlimited.

Prepared for:
Naval Electronic Systems Command
Washington, D. C. 20360

25 1 7

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral J. J. Ekelund
Superintendent

D. A. Schrady
Acting Provost

This investigation was sponsored by Mr. W. J. Dejka, NOSC, Code 8302. The work was performed by the author at the Naval Postgraduate School, Monterey, CA.

This report was prepared by:

D. E. Neil, Assistant Professor
Department of Operations Research

T. Andreason, Kapitaen Leutnant
Federal German Navy

Reviewed by:

K. T. Marshall, Chairman
Department of Operations Research

Released by:

William M. Tolles
Dean of Research

UNCLASSIFIED

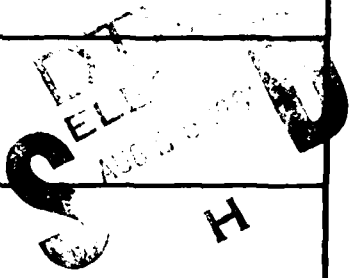
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: NPS55-81-004	2. GOVT ACCESSION NO.: AD-A103 280
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): EXAMINATION OF VOICE RECOGNITION SYSTEM TO FUNCTION IN A BILINGUAL MODE	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): D. E. Neil T. Andreason	8. CONTRACT OR GRANT NUMBER(s):
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School Monterey, CA 93940	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: N000398WR09041
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Postgraduate School Monterey, CA 93940	12. REPORT DATE: March 1981
	13. NUMBER OF PAGES: 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): Naval Electronic Systems Command Washington, D. C. 20360	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)



18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

VTAG
Voice Recognition
Automatic Word Recognition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

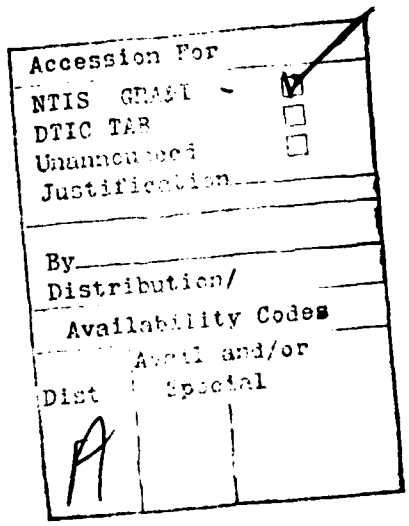
UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Foreword

This investigation was sponsored by Mr. W. J. Dejha, NOSC, Code 8302. The work was performed by the author at NPS, Monterey, CA.

This report is one of series concerned with the possible application of voice recognition technology in the military environment. It is the result of Professor Gary K. Poock's pursuit of the application of voice recognition in military systems and potential problem areas he has identified in the conduct of his efforts.



ABSTRACT

This report describes an experiment in which bilingual subjects (German/English) were used to examine the capability of Threshold Technology T600 voice recognition system to function in a bilingual mode.

Results suggested that the system functioned equally well in either language when training and testing was in one language. However, significant degradation was observed when training and testing was bilingual in nature.

I. INTRODUCTION

Traditionally man has interacted with machine through the use of his extremities (e.g., hands, feet, etc.) and reserved verbal behavior/speech for man-man communication. Recent technological advances in the design of speech recognition equipment, however, have suggested that this typical dichotomy of response modality is no longer absolutely necessary. The feasibility of employing speech as a man-machine control modality has been demonstrated in numerous research and applied efforts (Scott, 1978; Poock, 1980; Lea, 1980; Lea and Shoup, 1979; Doddington, 1980; Grady and Hicklin, 1976; Connolly, 1979; etc.).

In specific operational environments the possibility of using speech as a response mechanism capable of controlling machines possesses several potential advantages over traditional manual response systems. Lea (1980) and Martin and Welch (1980) have suggested that some of the advantage occuring to speech in a man-machine system are the result of the familiarity of speech as an output mechanism in most potential operators. Speech, as a result of the frequency and intensity of use is a "natural" and perhaps universal response system. As a result speech itself requires little in the way of training. Further, in situations wherein speech can be effectively used as an output mechanism in the interaction with machines, it may free the extremities and to some extent the decision making subsystems for functions incompatible with speech. The

net effect may well be an expansion of man's contribution in man-machine systems by taking full advantages of his capabilities.

Poock (1980) demonstrated the potential effectiveness of using speech as an input/control mechanism in a simulated Command-Control environment. Poock used voice recognition equipment to allow for verbal input to the ARPANET. His results indicated that voice input was faster than manual entry (17.5%); fewer errors were committed with voice than manual entry (183.2% more errors with manual); and information transfer was more efficient with voice than manual control (25.0% more information transcribed on a secondary task when using voice when compared to manual control). This was with operators who had only used voice input for 3 hours previously.

There are, of course, some problems associated with the use of speech as a control source in man-machine systems. Due to the nature of speech it is not private and therefore subject to unwanted monitoring. However, there are situations where it may be advantageous to hear an operator entering commands. One can hear what has been entered without having to ask or see what the operator has done. Further, it is sensitive to various ambient environmental influences, (e.g., noise, vibration, etc). Variability in speech as a result of native language, sex, age and perhaps physical condition or illness may influence speech output and subsequently the ability of speech recognition systems to function successfully. Obviously, manual control input systems are not without deficiencies and any application would need to examine various

strengths and weaknesses of both systems as well as environmental considerations and intended users.

The current effort selected one potentially degrading influence in speech recognition systems for study; namely "native" vs "official" language. In many military situations (e.g., NATO Command and Control Centers) it is possible for an operator to be required to interact with a system in an "official" language that is other than his/her "native" language. While the intended user may be quite fluent in the "official" language, the potential for reversion to his more natural vocal response or "native" language may be significant variable in system functioning. This tendency to revert to his more natural response may be fairly easily controlled during periods of routine or non-critical activity. However, such a tendency may increase with the intensity of activity or load placed on the operator. Such periods may be critical and intolerant of any influence which tends to degrade overall system functioning.

II. OBJECTIVE

The current effort was designed to examine the ability of a currently available voice recognition system to function in a bilingual mode. Specifically, could the Threshold Technology Inc., Model T600 discrete utterance voice recognition system be trained in two languages so that an utterance (i.e., an utterance consisting of a single word or continuous string of words not exceeding two seconds in duration) in either language would be recognized?

III. METHODOLOGY

Apparatus. Equipment consisted of a Model T600 Threshold Technology Inc., voice recognition system. The particular unit involved in the study was modified with the inclusion of additional memory modules providing for up to 256 .1 to 2 second discrete utterances. In the experiment 105 discrete utternaces were used. Appendix A contains the 105 utterance list.

In the actual experiment the T600 unit was placed in an Industrial Acoustic Co., Inc. sound attenuating booth. The purpose of conducting testing in a controlled ambient noise environment was to minimize acoustic influences as well as other environmental influences which may impair voice recognition system performance, as well as providing distracting stimuli to subjects.

Subjects. Subjects consisted of 12 males and four females. Male subjects were German officer students at the Naval Postgraduate School. Female subjects were wives of German students attending the Naval Postgraduate School. All subjects were bilingual (German/English) with German being the "native" language in each case. All subjects were volunteers and received no compensation for their participation. Subjects' ages ranged from 26-37 years.

Procedure. A 105 utterance list was prepared for use in the study. Utterances were selected on their possible application in Command-Control type environment. No attempt

was made to control for syllable count in either language, nor were any utterances accepted or rejected on the basis of their potential for enhancing recognition.

The T600 requires that each subject "train" each utterance a total of 10 times. That is, a subject must repeat each utterance 10 times in order to provide a basis for comparison in the testing mode. In the present experiment subjects were required to "train" the system with the utterance list three times. Subjects repeated each utterance 10 times in English for the test of recognition with training and testing in English; repeated each utterance 10 times in German for the German training and testing portion of the experiment; and repeated each word 5 times in German and 5 times in English for the combined English/German portion of the study.

Therefore, subjects trained the system under each of the three conditions followed by testing on that condition, then proceeded to the next condition, etc. In the mixed condition subjects trained and tested each utterance in both English and German.

It should be mentioned that translation from English to German was accomplished by one of the experimenters to provide a standard German utterance list as well as a standard English utterance list. This was done to reduce variability in the utterance list for German. It was observed that without such a standardization procedure considerable variability in translation of English to German was possible.

The order of language or conditions a subject received was randomized to prevent the possible interaction of training sequence with system performance.

Performance measures. Performance was considered in terms of recognition accuracy under the training/testing conditions described above. Misrecognition (i.e., incorrect recognitions of an utterance) and inability of the voice recognition system to match the test utterance with any trained utterance (signaled by an auditory "beep" from the T600) were considered as errors and given equal weight in the analysis.

Experimental design. The interest was obviously whether a significant difference existed between the three training conditions previously described and voice recognition system performance. The design selected involved repeated measures in which each subject served as his own control and was therefore tested under each training conditions. This particular design was selected as a result of the limited number of subjects available, and the ability of the design to isolate training effect variability and reduce variability associated with individual differences (Myers, 1967; Weiner, 1962). That is, repeated measures method should provide some control for differences between subjects.

In addition, due to the nature of the data, analysis was performed on raw data and on transformed data. An arcsin transformation was used to put the data into a form that would most nearly satisfy the assumptions underlying analysis of variance (Weiner, 1962).

IV. RESULTS AND DISCUSSION

Table I presents a summary of misrecognition/non-recognition errors of voice recognition equipment under the training/testing conditions used. Table I suggests that overall system performance was degraded under the mixed training/testing conditions when compared with either English or German alone. Further, performance with the subject's "native" language (i.e. German) would appear to be slightly superior to the performance in the secondary language (i.e., English).

Table II presents the results of analysis of variance using raw data. Analysis suggested that between subject variability was not highly significant. It should be remembered that the design selected should reduce individual subject variability and therefore provide some measure of control for differences between subjects.

Within subject variability was observed to be statistically significant ($p<.01$). This would suggest that within individual subject performance under the various language conditions was highly variable.

Conditions or language used during training was observed to be highly significant ($p<.001$). This finding suggests that in the raw data, at least, training conditions impacted significantly on voice recognition performance.

Table III presents a similar analysis on the data following an arcsin transformation. Transformed data supported analysis on raw data in that a significant within subject variation was observed ($p<.01$) and a highly significant training

condition effect ($p<.001$). Like the raw data, analysis of transformed data suggested a potentially significant between subject variation. Granted the degree of statistical significance ($p<.05$) was somewhat lower than the within or training condition sources of variation, the implication is that a possible between subject influence was present. This finding may be particularly interesting in view of the experimental design employed.

The analysis on both raw data and arcsin transformed data both suggest a highly significant condition or training language effect. Obviously, it would be necessary to attempt to determine the nature of the training influence. A Newman-Keuls test on the difference between all possible pairs of treatment was conducted in an attempt to examine the dominant training influence. Treatment totals were used rather than treatment means in the Newman-Keuls analysis as a result of the number of observations under each treatment or training condition being equal (Weiner, 1962).

Newman-Keuls analysis of raw data suggested no difference between training/testing in English and training/testing in German. Therefore, the slight improvement in performance of German over English suggested in Table I was not statistically significant. However, analysis of the difference between system performance using English alone when compared to the mixed English/German was significant ($p<.01$). Furthermore, German alone when compared to the mixed English/German training/testing condition was also highly significant ($p<.01$).

TABLE I. SUMMARY OF ERRORS UNDER THE ENGLISH, GERMAN AND ENGLISH/GERMAN CONDITION

	English (trng/testg)	German (trng/testg)	English (mixed trng/testg)	German (mixed trng/testg)	Combined (mixed trng/testg)
Errors	124	78	400	334	734

TABLE II. ANALYSIS OF VARIANCE USING RAW DATA

Source of Variation	ss	df	ms	F
Between subjects	1300	15	86.6	1.86 NS***
Within subjects	18144	32	567	12.32*
Training language	16761.5	2	8380.7	182.18*
Residual	1382.5	30	46	

*P < .01

TABLE III. ANALYSIS OF VARIANCE USING ARCSIN TRANSFORMED DATA

Source of Variation	ss	df	ms	F
Between subjects	.31	15	.021	2.1*
Within subjects	3.11	32	.097	9.7**
Training language	2.8	2	1.4	140**
Residual	.31	30	.010	

*P < .05
**P < .01
*** < .10

Therefore, in the raw data case, it would appear that voice recognition with either of the test languages was roughly equivalent (i.e. no statistically significant difference between German and English). However, recognition performance was severely degraded when the two languages were combined.

Analysis using the Newman-Keuls procedure on transformed data yielded results similar to the raw data. Analysis revealed no statistically significant differences between English and German when training/testing involved single language conditions. However, as when raw data was analyzed, a statically significant difference in system performance was observed when English alone was compared to mixed English/ German ($p < .01$) and when German alone was compared to mixed English/German ($p < .01$).

In an attempt to determine whether one language contributed a disproportionate amount of performance degradation under the mixed language condition, an analysis of performance of English and German in the combined test was conducted. That is, recognition errors in English and recognition errors in German in the combined training situation were evaluated to determine the contribution of each to overall performance degradation.

The Newman-Keuls procedure was used to examine treatment totals under the two conditions. The results indicated no statistically significant difference between the languages in testing. Therefore, it would appear that neither language

was primarily responsible for the reduction of recognition performance during testing.

As mentioned earlier, subject population included four females. Due to the small number of females, statistical analysis was not considered. Figure 1, does present a graphical representation of the average performance of male subjects as compared to female subjects. The figure suggests that recognition performance using females was slightly inferior for either German or English while recognition performance with females under the mixed condition was slightly superior to that using male subjects.

There are a number of potentially important variables which may partially explain the results suggested in Figure 1. First, as already suggested the fact that the sample consisted of 12 males and four females renders any attempt to consider sexual differences questionable at best. Further, male subjects were all students at the Naval Postgraduate School and were therefore probably more accustomed to functioning in an environment requiring the use of English. In addition, as a result of their student status they were more familiar with the testing environment and the research process. Male subjects were, therefore, probably more "comfortable" in the experimental situation. All of the above factors probably contributed to observe differences between males and females.

In summary, the results of the present effort suggest no difference between the languages used here when both training and testing were restricted to a single language. However,

recognition performance was significantly degraded when the system was trained to respond in either language.

The results are not surprising when one considers the manner in which the T600 system operates. The process employed by the system involves the extraction of a matrix of distinctive speaker characteristics for each repetition of an utterance. At the conclusion of the 10 training passes for each utterance a single reference matrix is formed which contains the dominant characteristics of each utterance. During testing, an utterance is compared to the reference matrix in an attempt to determine whether the utterance matches a trained utterance.

In the bilingual mode it can be postulated that substantial variation was associated with each utterance. Such a situation would provide an extremely complex array increasing the difficulty of the T600 system to accurately develop a reference matrix. Therefore, it can be suggested that reference matrices lacked the definition necessary for desired accuracy.

Conclusion

Based on the results of the present study it would appear that other T600 is quite capable of functioning with either English or German but not the two in combination. Therefore, it does not appear to be a viable input instrument in situations which may involve the potential bilingual presentation of commands. Granted in most situations the instrument

would not be required to function under such conditions. Further, given user awareness of the inability to function in a bilingual mode, procedural controls could be developed which would minimize the potential ramifications of the T600 inability to recognize two dissimilar languages.

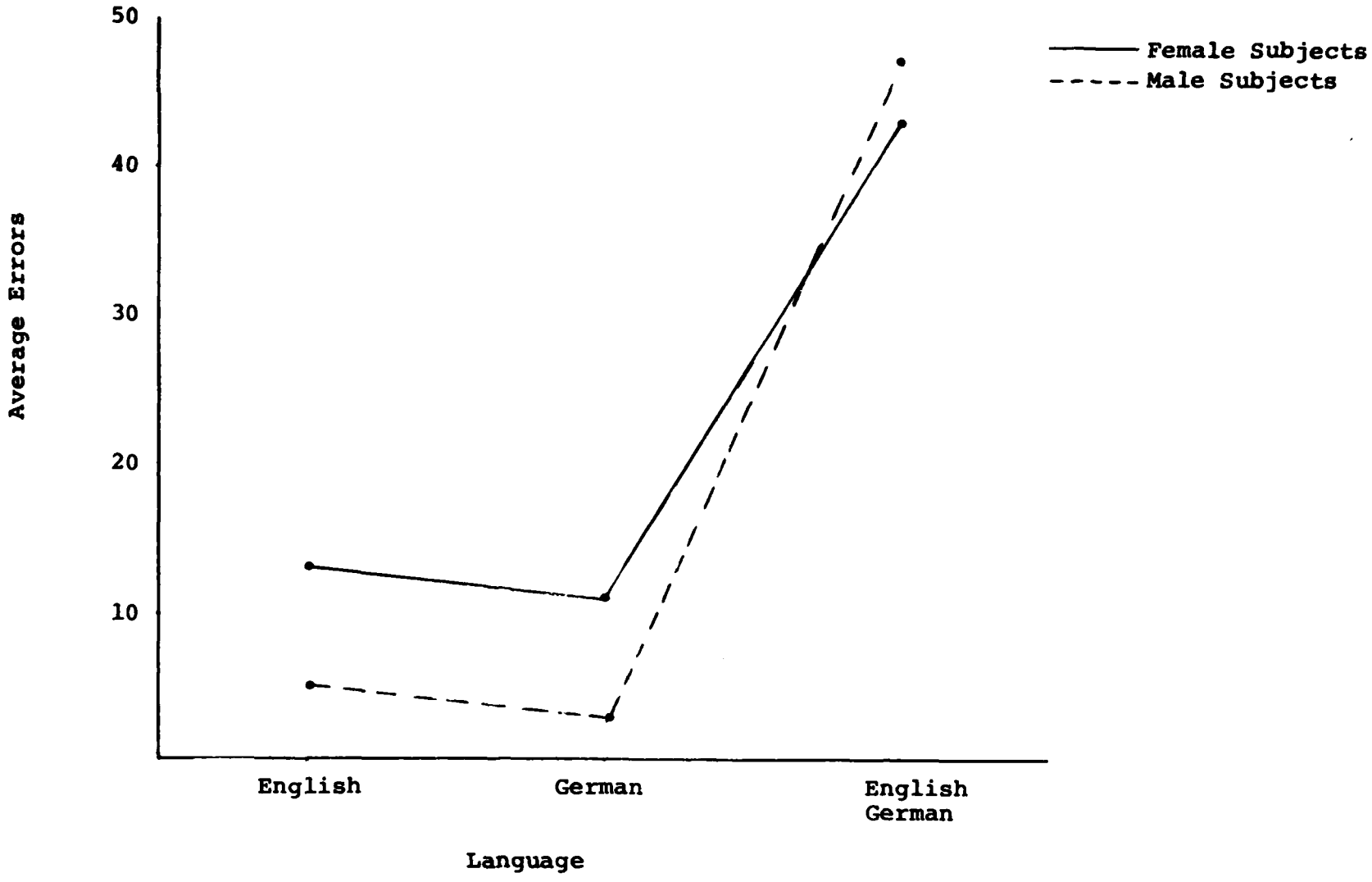


Figure 1. Performance differences for Males and Females under language conditions.

REFERENCES

Connally, D. W., Voice Data Entry in Air Traffic Control. National Aviation Facilities Experimental Center, Dept. #FAA-NA-79-20, August 1979.

Doddington, G., Voice Identification for Entry Control. Paper presented at Voice Interactive Systems: Applications and Payoffs Meeting, Dallas, Texas, May 13-15, 1979.

Grady, M. W. and Hicklin, M., Use of Computer Speech Understanding in Training: A Demonstration System for the Ground Controlled Approach Controller. Naval Training Equipment Centers, Dept. #74-C-0048-1 (A.D.-A033 327), December 1976.

Lea, W. A., The Value of Speech Recognition Systems In: Trends in Speech Recognition, W. A. Lea (Ed). Englewood Cliffs, NJ: Prentice Hall, 1980.

Lea, W. A. and Shoup, J. E., Review of ARPA SUR Project and Survey of Current Technology in Speech Understanding. ONR Speech Community Laboratory Dept., 1979.

Martin, T. B. and Welch, J. R., Practical Speech Recognizers and Some Performance Effectiveness Parameters. In: Trends in Speech Recognition. W. A. Lea (Ed). Englewood Cliffs, NJ: Prentice Hall, 1980.

Myers, J. L., Fundamentals of Experinental Design. Boston: Allyn and Bacon, 1966.

Poock, G. K., Experiments with Voice Input for Command and Control. Naval Postgraduate School Technical Report NPS-55-80-016, April 1980.

Scott, P. B., Word Recognition. Rome Air Development Center Report TR-78-209, (AD-A061-545), September 1978.

Weiner, B. J., Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

APPENDIX A

0. one	25. delay	50. minutes
1. two	26. designate	51. name
2. three	27. distance	52. neutral
3. four	28. dive	53. north
4. five	29. drop	54. now
5. six	30. east	55. existing
6. seven	31. end	56. off
7. eight	32. enemy	57. on
8. nine	33. envelope	58. contact
9. zero	34. execute	59. detect
10. air	35. fix	60. mission
11. status	36. fire	61. orders
12. altitude	37. forces	62. others
13. at	38. friendly	63. own
14. attack	39. patrol	64. pass
15. heading	40. event	65. sortie
16. barrier	41. help	66. circle
17. bearing	42. if attacked	67. marker
18. azimuth	43. label	68. update
19. cancel	44. launch	69. plot
20. new	45. lay barrier	70. point
21. course	46. list	71. position
22. speed	47. maneuver	72. probability
23. cover	48. map	73. proceed
24. degrees	49. minefield	74. refuel

75. report
76. self
77. sensor
78. south
79. space
80. missile
81. station
82. submarine
83. surface
84. target
85. time
86. track
87. unknown
88. west
89. aircraft
90. radar
91. sonar
92. sonobuoy
93. range
94. latitude
95. longitude
96. vector
97. remote
98. distress
99. bomb
100. weapon
101. fly to
102. torpedo
103. predict
104. base

DISTRIBUTION LIST

	No. of Copies
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	2
Dean of Research Code 012 Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
Asst. Prof. D. E. Neil Code 55Ni Naval Postgraduate School Monterey, CA 93940	183